International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC)
ISSN(P): 2250-1568; ISSN(E): 2278-9448
Vol. 4, Issue 1, Feb 2014, 39-44

A SURVEY ON: SENTIMENT ANALYSIS FOR MOVIE DOMAIN IN MOBILE ENVIRONMENT

SAVITA LAXMAN HARER & YOGESH P. SAYAJI

Dr. D. Y. Patil College of Engineering, Ambi, Pune, Maharashtra, India

ABSTRACT

Many people express their views and opinions about products and service In this paper we focus on specific domain that is movie reviews. This literature survey paper describes importance of opinion mining, classification method for sentiment analysis that are Naive Bayes, Maximum entropy, Support Vector machine, task of opinion mining at different levels, feature based summarization of movie reviews. We propose a novel approach based on LSA method for product feature identification. We consider here sentiment classification accuracy and system response time to design the system in mobile environment and movie rating is based on sentiment classification result.

KEYWORDS: Sentiment Analysis, Sentiment Classification, Natural Language Processing, Machine Learning, Opinion Mining, Emotions

INTRODUCTION

Sentiment analysis has various different names as like opinion mining, sentiment mining, subjectivity analysis, affect analysis, motion detection, opinion spam detection etc. Generally, sentiment analysis is a type of natural language processing which determine the attitude, feelings or emotions, opinions with respect to specific topic and product or services. In earlier days when we want to purchase any product we used to ask our friends and family those who have knowledge about that product. But now a days there is a rapid development of World Wide Web technology consisting of many web sites such as for products (www.amozon.com), movies (www.imdb.com), hotels (www.tripadvisor.com), and restaurants (www.yelp.com) that motivates people to convey their opinions about specific topic in which they are interested. Opinions are key influencers of behaviors. The text information specifies facts and opinions. The facts are objective expression and opinions are subjective expression. In short we say that opinion mining [1] is an automated extraction of subjective content from text and identifying the orientation such as positive and negative in that text.

Sentiment analysis is more widely used in industries. For example, it is beneficial for the business for selling their products and manufacturer can collect the feedbacks from the customers through web technology about that products. As the numbers of customers are increasing reviews received by organizations are also growing in large amount. When person want to purchase any product he/she first reads the reviews/ opinions of other people's about that product and helpful for them in decision making process whether to purchase specific product or not? Thus reviews are helpful for both the producers and consumers. The words sentiment and opinion can be used interchangeably. Both are widely used in academia.

As today mobile phones have become most important part of our life .As we already know there are lots of applications of mobile phones. In mobile environment it is inappropriate to display detailed review because of small screen size [2]. Thus we propose review summarization mechanism to reduce the size of review information. The system will summarize the movie reviews into positive review classes and negative review classes and provide the users an overview about that review. Thus, automatic opinion mining and summarizing has gained importance in research field.

OPINION MINING AT DIFFERENT LEVELS

Classification of Opinion Mining at Document Level

Document level opinion mining is classifying each review document into positive, negative and neutral classes.

Classification of Opinion Mining at Sentence Level

Sentence level opinion mining is identifying the given sentence is subjective or objective and after that find out opinion of subjective sentence as positive, negative and neutral.

Classification of Opinion Mining at Feature Level

The function of opinion mining at feature level is to extracting the features of commented object and after that determine the opinions on the features are positive, negative and neutral.

SENTIMENT CLASSIFICATION

There are various machine learning methods for automatic sentiment classification that are Naïve Bayes, Maximum Entropy and Support Vector Machines.

Table 1: Comparison of Machine Learning Technique Accuracies Based on Number of Features

Sr. No.	Author	Features	No. of Features	Frequency or Presence	Dataset	Classifier with Performance Accuracies		
1	Liu, Lu and Jou(2012)	Unigrams	36084	Presence	IMDB	86.5(SVM)		
		Unigrams with occurrences more than 3	15026	Frequency as filtering criterion		86.25(SVM)		
		Unigrams using the frequency criterion based on 3	861	Frequency criterion		81.2(SVM)		
2	Pang and Lee(2002)					*NB*	*ME*	*SVM*
		Unigrams	16165	Frequency	IMDB	78.7	N/A	72.8
		Unigrams	16165	Presence		81.0	80.4	82.9
		Unigrams +Bigrams	32330	Presence		80.06	80.08	82.7
		Bigrams	16165	Presence		77.3	77.4	77.1
		Unigrams + POS	16695	Presence		81.5	80.4	81.9
		Adjectives	2633	Presence		77.0	77.7	75.1
		Total 2633 unigrams	2633	Presence		80.3	81.0	81.4
		Unigrams + position	22430	Presence		81.0	80.1	81.6

As shown in Table1 named “Comparison of Machine learning technique accuracies based on number of features” Pang and Lee [3, 4] compared the performance of Naive Bayes, Maximum Entropy and Support Vector Machines on different features.

Following Result is observed from Table 1

- Feature presence is more important than feature frequency.
- Accuracy of classifier improves if all frequently occurring words from all parts of speech are taken.
- Classifier accuracy increases by position information.
- When the numbers of features are less at that time Naïve Bayes perform better than SVM. When the numbers of features are increasing SVM performs better.
- The accuracy of classifier falls by Bigrams.

Table 1 also shows that Liu, Lu and Jou [2] the performance of SVM classifier with respect to numbers of features. When the numbers of features are large unigrams as features out performs than other. The second one is reduces the numbers

of features and accuracy is same and third one is that numbers of features are less with accuracy 81.2%.

Table 2: Summary of Classification Technique

Sr. No.	Author	Feature	Dataset	Classifier	Accuracy
1	Kaiquan Xu(2011)	Linguistic feature	Amazon reviews	Multiclass SVM	61%
2	Xue Bai(2011)	Information gain	Movie reviews	Naive Bayes	92%
3	Yulan He(2010)	Self trained	Movie reviews	Lexical approach	74.7%
4	Zhu Jian(2010)	Odds ratio	Movie reviews	Back propogation	86%
5	Rudy(2009)	Document frequency	Movie reviews	SVM,Hybrid	89%
6	Qingliang Miao(2009)	POS,Apriori	Amazon reviews	Lexical approach	87.6%
7	Kennedy and Inkpen(2006)	frequencies	Movie review	SVM	86.2%
8	Gamon(2005)	Stemmed terms	Car reviews	Naive Bayes	86%
9	Bai(2005)	Dependence among words	Movie reviews	Two-stage Markov Blanket	87.5%
10	Pang and Lee(2004)	Based on minimum cuts	Movie reviews	Naive Bayes	86.4%

As shown in Table2 named “Summary of classification technique” shows large amount of work has been done on Product reviews and movie reviews. In this paper we focus on Movie reviews. The movie reviews different from product reviews [5] the reason behind is that when person writes movie review he/she comments on movie elements like music, dialogues as well as movie related people like director, screenwriter, actor, actress etc. On the other hand there are specific commented features in product reviews because people may like some features and dislike other which one is a difficult to classify opinion orientation of review as positive or negative. Feature specific reviews occur less often in movie reviews. Another reason is that there are a lot of comparative sentences in product reviews and people discuss about other product in reviews. So movie review mining is more challenging application that other types of review mining.

As shown in Table 3 named “Summary of algorithms for sentiment analysis”.

Table 3: Summary of Algorithms for Sentiment Analysis

Sr. No	Author	Algorithm	Dataset	Feature
1	Khairnar and Kinikar(2013)	LSA(Latent Semantic Analysis)	Mobile reviews	Identification of mobile feature
2	Liu , Hsaio, Lee, Lu and Jou (2012)	LSA(Latent Semantic Analysis)	Movie reviews	Movie Feature Identification
3	Liu , Hsaio, Lee, Lu and Jou (2012)	Frequency based		Movie Feature Identification
4	Liu , Hsaio, Lee, Lu and Jou (2012)	LSA-based filtering approach		Feature based summarization
5	Yu, Liu, Huang and An(2012)	S-PLSA(Sentiment Probabilistic Latent semantic analysis))	Movie reviews	Provide probabilistic framework to analyze sentiments in reviews
6	Zhu Zang (2008)	SVR(Support vector regression algorithm)	Product reviews	Provide radial-basis kernel function(RBI) for optioning structural patterns in data
7	T.Hofman (2001)	PLSA(Probabilistic Latent semantic analysis)	IMDB	Movie Feature Identification
8	V. Hastzivassiloglour(1997)	PMI-IR	Chinese product reviews	Provide semantic orientation of a phrase

In this paper we focus on feature based summarization of movie reviews [2]. The feature based summarization consists of product features on which the customers have expressed their opinions and an opinion word means that opinion

information about that product. So for feature based summarization product feature identification and opinion word identification are important. For opinion word identification we taking Parts of speech approach because they provide a crude form of word sense disambiguation.

CONCLUSIONS

As we knows that cellular phones are very essential for everyone. But digital content displayed in mobile phones are limited due to the small size of screen. From Table1 and Table 2 we observe that Support Vector Machine classifier is best for sentiment classification of movie reviews according to performance accuracy. SVM classifier classifies the movie reviews into positive and negative review classes and movie rating information depends on sentiment classification result. From Table1 we observe that the system using unigrams with presence feature will have 40 462 features it takes about 120s for loading classification model. But, it is infeasible on mobile platform. So for reducing the number of features we propose frequency as filtering criterion. If we use the frequency criterion then numbers of features are reduced to 1902 and it takes about 6s for loading classification model and it is feasible on mobile platform. Hence we propose an LSA based filtering mechanism which allows the user to determine the feature in which he/she is interested. From Table3 we propose Latent Semantic Analysis algorithm for identification of movie feature. We compare the LSA algorithm with PLSA algorithm. In future it extend to other product domain easily. In future our work can extend to achieve greater efficiency in LSA method.

REFERENCES

1. G. Jaganadh 2012. Opinion mining and Sentiment analysis CSI communication.
2. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, "Movie Rating and Review Summarization in Mobile Environment", IEEE VOL. 42, NO. 3, MAY 2012
3. Pang, Bo and Lee, Lillian and Vaithyanathan, Shivakumar, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of EMNLP 2002.
4. Pang, Bo and Lee, Lillian and Vaithyanathan, Shivakumar, Thumbs up?: Sentiment Classification using machine learning techniques, In Proceedings of the ACL-02 conference on Empirical Methods in Natural Language, 2002.
5. L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in Proc. 15th ACM Int. Conf. Inf. Knowl. Manage., 2006, pp. 43–50.
6. Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for CompetitiveIntelligence", Decision Support Systems 50 (2011) 743–754.
7. ZHU Jian, XU Chen, WANG Han-shi, "" Sentiment classification using the theory of ANNs", The Journal of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 58–62 .
8. Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach.", Journal of Informetrics 3 (2009) 143–157.
9. Qingliang Miao, Qiudan Li, Ruwei Dai, "AMAZING: A sentiment mining and retrieval system", Expert Systems with Applications 36 (2009) 7192–7198.
10. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Computational Intelligence, vol. 22, pp. 110–125, 2006.

11. T. Hofmann, "Probabilistic Latent Semantic Indexing," Proc. 22ndAnn. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 50-57, 1999.
12. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc.10th ACMSIGKDD Int. Conf. Knowl. Discov. Data Mining, 2004, pp. 168–177.
13. V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in Proc. 8th Conf. Eur. Chap. Assoc. Comput. Linguist., Morristown, NJ: Assoc. Comput. Linguist., 1997, pp. 174–181.
14. T. Joachims, Learning to Classify Text Using Support Vector Machines:Methods, Theory and Algorithms. Norwell, MA: Kluwer, 2002.
15. T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis,"Mach. Learn., vol. 42, no. 1/2, pp. 177–196, 2001.
16. T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in Proc. EMNLP, 2004, pp. 412–418.
17. P.Turney 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceeding of Association for Computational Linguistics, pp. 417--424.